Random Variables and Probability Distributions **Notes 02**

Associated Reading: Wackerly 7, Chapter 1, Section 3, Chapter 2, Section 11, Chapter 3, Sections 1-3 and 11, and Chapter 4, Sections 1-3

We'll start with a motivating example. We flip a fair coin three times. What may we observe?

Sample space $S = \{HHH, HHT, HTH, THH \dots TTT\}$. $|S| = 8$.

From this information, we can compute, e.g., the probability of observing exactly one tail, T, via the sample-point method. And we can, laboriously, generate a table of probabilities:

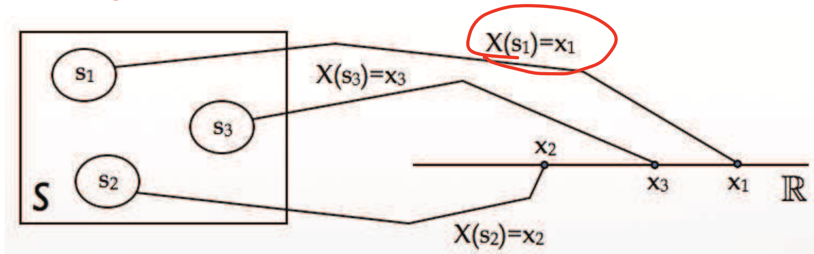
simple events with prob.

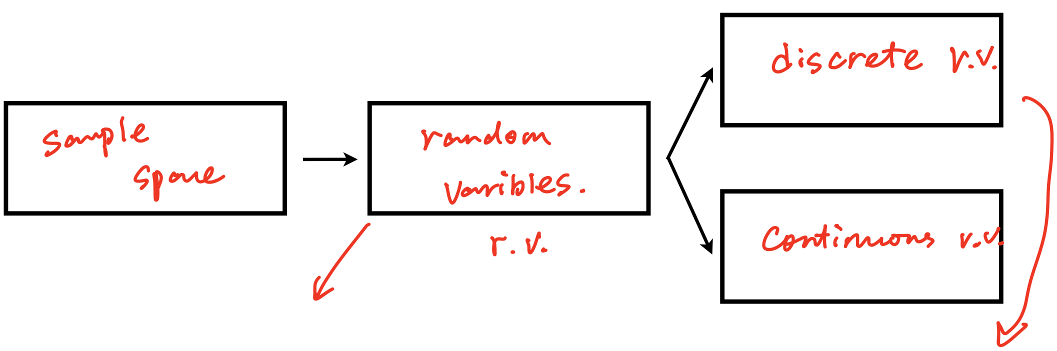
$$\begin{aligned} P[\text{"no tails"}] &= P[\text{HHH}] \\ \frac{3}{8} = \boxed{P[\text{"one tail"}]} &= P[\text{HHT} \cup \text{HTH} \cup \text{THH}] \\ P[\text{"two tails"}] &= P[\text{HTT} \cup \text{THT} \cup \text{TTH}] \\ P[\text{"three tails"}] &= P[\text{TTT}] \end{aligned}$$

Capital: X, Y, Z

There's a better way to portray this information: using a *random variable*, a function that maps events in a sample space $\mathcal{S}$ to $\mathbb{R}^n$ (where $\mathbb{R}^1$ is the real number line).

$\mathbb{R}^1$: real number.
$\mathbb{R}^n$: n-dimensional vector.





different r.vs are allowed.

This refers to the possible values of the r.v.

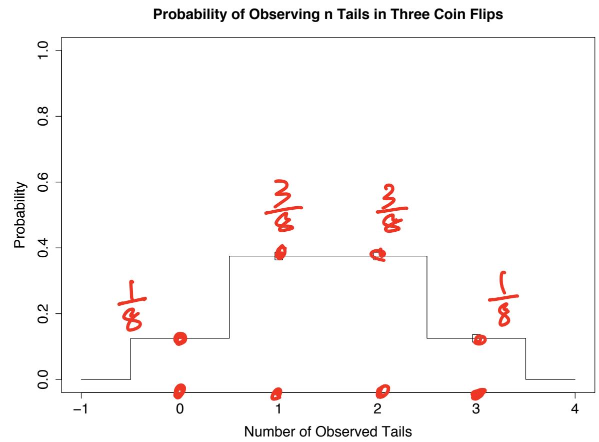
e.g. 1, 2, 3 ⋯ ← discrete r.v.

[0, 1] ← continuous r.v.

P(Y(S) = 1) ⋯

$\boxed{P(Y(s)=0)} = P(\{s \in S, \text{ st: } Y(s)=0\}) = P(\{HHH\}) = \frac{1}{8}$

A



There are no specific rules dictating how one defines the set of possible values that the random variable $Y(s)$ can take on for a given experiment. Most often, the choice is intuitively obvious: for instance, if Y represents the number of observed tails in three flips, it should take on the values $\{0,1,2,3\}$, as shown at right.

Some points to make about random variables:

- Ultimately, in the big picture, why should I care? After all, I still have to define the sample space in order to determine $P[Y(s) = y]$, right?

  r.vs provide us a unified framework to work with prob.

- Writing $P[Y(s) = y]$ is laborious. Can I write $P[Y = y]$ instead? How about $P[Y]$? Or $P[y]$?

  $P(Y=y)$ ok ✓ $\quad P(Y)$. ✗ $\quad P(y)$ ✗ $\quad p(y) := P(Y=y)$. is allowed.

  e.g. $f(x)=x^2$. Y is r.v. $f(Y)=Y^2$ is also r.v.

- *Functions of random variables are themselves random variables!*

Starting here, we will examine the material at the beginnings of Chapter 3 and Chapter 4 concurrently, since presenting definitions associated with discrete and continuous random variables side-by-side also allows us to more easily highlight any differences between them.

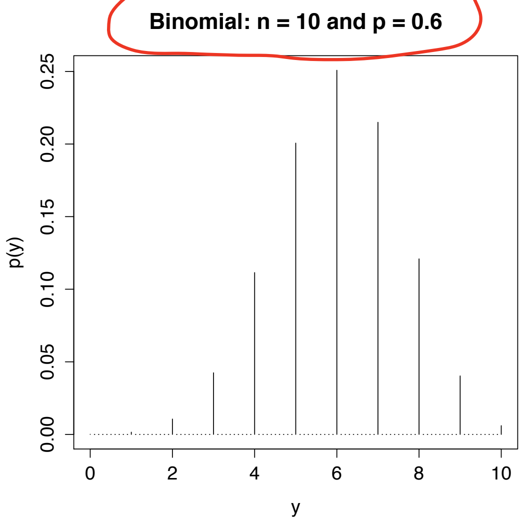
DISTRIBUTIONS:	**DISCRETE**	**CONTINUOUS**
Name	probability mass function (pmf)	probability density func. (pdf).
Symbol	$p(y) := P(Y=y)$.	$f(y)$
Fundamental Properties (Theorems 3.1/4.2)	$p(y) \in [0,1]$	$f(y) \in [0,+\infty)$.
	$\sum_{\text{possible value of } y} p(y) = 1$	$\int_{\text{domain of } f(y)} f(y)\,dy = 1$
	$P(a \le Y \le b) = \sum_{y\in[a,b]} p(y)$	$P(a \le Y \le b) = \int_a^b f(y)\,dy$

$P(Y=a) = 0$

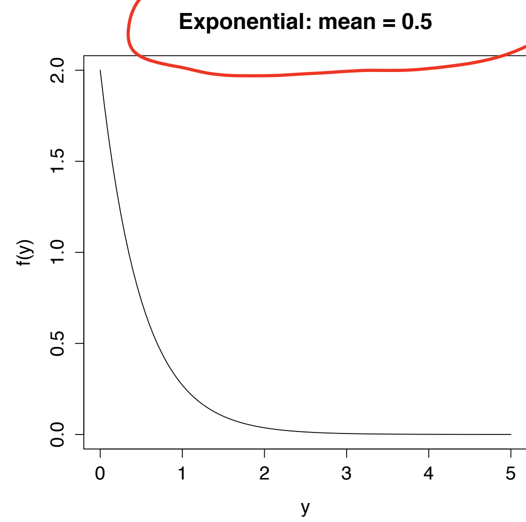
$= P(a \le Y \le a)$

$= \int_a^a f(y)\,dy = 0$ 0

pmf



Discrete distribution



Continuous distribution

→ **EXAMPLE.** Wackerly 7, Exercise 3.3

3.3 A group of four components is known to contain two defectives. An inspector tests the components one at a time until the two defectives are located. Once she locates the two defectives, she stops testing, but the second defective is tested to ensure accuracy. Let Y denote the number of the test on which the second defective is found. Find the probability distribution for Y.

pmf

let D = "defective"

$S = \{ \underbrace{DD}_{Y=2},\ \underbrace{\bar{D}DD,\ D\bar{D}D}_{Y=3},\ \underbrace{\bar{D}\bar{D}DD,\ \bar{D}D\bar{D}D,\ D\bar{D}\bar{D}D}_{Y=4} \}.$

pmf.

y	2	3	4
$p(y)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

$p(2)$ $p(3)$

$p(y) = 0$ $y \ne 2,3,4$

→ **EXAMPLE.** What is c, given the following pdf?

$$f(y) = \begin{cases} c\exp\left(-\frac{y}{2}\right) & y \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

$\exp(x) = e^x$

relevant property: $\int_{\text{domain of } y} f(y)\,dy = 1.$

$\int_0^\infty c \cdot \exp(-\frac{y}{2})\,dy = 1 \implies$ solve for c.

$c \cdot \int_0^\infty \exp(-\frac{y}{2})\,dy.$

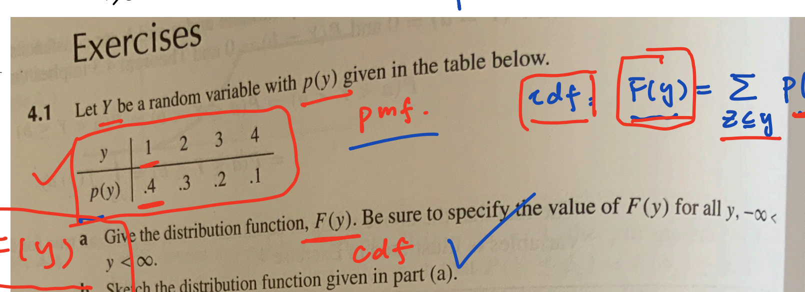
$c \cdot \left(-2\exp(-\frac{y}{2})\Big|_0^\infty\right) = c \cdot [0 - (-2)] = 2c.$

$\Downarrow\ c = 1/2.$

Now we will lay out the definitions associated with the *cumulative distribution function*, or cdf:[a]

CDFs:	**DISCRETE**	**CONTINUOUS**
Symbol	diff $F(y^-) \to F(y)$	$F(y)$: cdf.
Definition	$F(y) := \sum_{\text{any } z \le y} p(z)$	$F(y) := \int_{z \le y} f(z)\,dz$
Limiting Properties (Theorem 4.1)	$F(-\infty) = 0$; $F(\infty) = 1$; $F(\infty) = \sum_{z \le \infty} p(z) = 1$	if $y_1 \le y_2$ $F(y_1) \le F(y_2)$ ($p(z) \ge 0$, $f(z) \ge 0$).
Relationship to pmf/pdf	If $F(y)$ steps up in value by amount x. then $p(y) = x$.	$f(y) = \frac{d}{dy} F(y)$
~~**Relationship to Percentile p**~~	later.	
Relationship to Summed/Integrated Probability (Theorem 4.3)	$P(a < Y \le b) = F(b) - F(a)$. ⚠ $P(a < Y \le b) \ne P(a \le Y \le b)$ $\propto$ $P(a) \ne 0$.	$P(a < Y \le b) = F(b) - F(a)$. ⚠ $P(y) = 0$ $P(a < Y \le b) = P(a \le Y \le b)$.

→ **EXAMPLE.** Wackerly 7, Exercise 4.1


Exercises

4.1 Let Y be a random variable with $p(y)$ given in the table below.

y	1	2	3	4
$p(y)$	.4	.3	.2	.1

a Give the distribution function, $F(y)$. Be sure to specify the value of $F(y)$ for all y, $-\infty < y < \infty$.

b Sketch the distribution function given in part (a).

pmf. cdf

cdf: $F(y) = \sum_{z \le y} p(z)$

a).

for $y < 1$ $F(y) = 0 = \sum_{z < y < 1} p(z)$

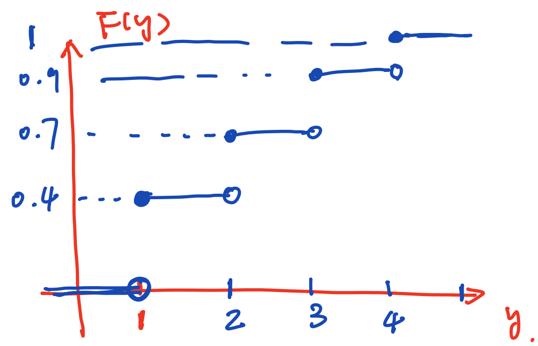
for $1 \le y < 2$ $F(y) = 0.4 = p(1)$.

for $2 \le y < 3$ $F(y) = 0.7 = p(1) + p(2)$

$3 \le y < 4$ $F(y) = 0.9$

$y \ge 4$ $F(y) = 1 = p(1) + p(2) + p(3) + p(4)$.

y	$F(y)$
$y < 1$	0
$1 \le y < 2$	0.4
$2 \le y < 3$	0.7
$3 \le y < 4$	0.9
$y \ge 4$	1



$y = 1$ $F(1) = 0.4$

[a]In Wackerly 7 (and some other texts), this is referred to as simply the "distribution function" or "df."

→ **EXAMPLE.** Wackerly 7, Exercise 4.17 (skipping c)

4.17 The length of time required by students to complete a one-hour exam is a random variable with a density function given by

Y

$$f(y) = \begin{cases} cy^2 + y, & \boxed{0 \le y \le 1,} \\ 0, & \text{elsewhere.} \end{cases}$$

a Find c.
b Find $F(y)$.
~~**c** Graph $f(y)$ and $F(y)$.~~
d Use $F(y)$ in part (b) to find $F(-1)$, $F(0)$, and $F(1)$.
e Find the probability that a randomly selected student will finish in less than half an hour.
f Given that a particular student needs at least 15 minutes to complete the exam, find the probability that she will require at least 30 minutes to finish.

a). property: $\int_{\text{domain}} f(y)\,dy = 1$

$\Downarrow$

$\int_0^1 (cy^2 + y)\,dy = 1.$

$\Downarrow$

$c \cdot \int_0^1 y^2\,dy + \int_0^1 y\,dy = 1$

$c \cdot \frac{y^3}{3}\Big|_0^1 + \frac{y^2}{2}\Big|_0^1 = 1$

$\boxed{\frac{1}{3}c + \frac{1}{2} = 1 \Rightarrow c = \frac{3}{2}}$

b). $F(y) = \boxed{\int_0^y f(z)\,dz,}$

$= \int_0^y \cdot (\frac{3}{2}z^2 + z)\,dz.$

$= \frac{1}{2}y^3 + \frac{1}{2}y^2. \ (0 \le y \le 1)$

y	$y<0$	$[0,1]$	$y \ge 1$
$F(y)$	0	$\frac{1}{2}y^3 + \frac{1}{2}y^2$	1

d). $F(-1) = 0$

$F(0) = 0$

$F(1) = 1$

$\boxed{P(Y \ge a) = P(Y > a)}$

e). $P(Y < \frac{1}{2}).$

$= \int_0^{\frac{1}{2}} f(y)\,dy$

$= F(\frac{1}{2}) - F(0).$

$= \frac{3}{16}$

f). $P(Y \ge \frac{1}{2} \mid Y \ge \frac{1}{4}).$

$= \frac{P(Y \ge \frac{1}{2} \cap Y \ge \frac{1}{4})}{P(Y \ge \frac{1}{4})}$

$= \frac{P(Y \ge \frac{1}{2})}{P(Y \ge \frac{1}{4})} = \frac{104}{123}$

$\ne F(\frac{1}{2})$

And now we lay out the definitions associated with the *expected value* operator. Applied to the random variable Y, it gives the mean (or first *moment*) of the distribution from which Y is sampled.

	DISCRETE	**CONTINUOUS**
Symbol	$E[Y]$ (or μ)	
Definition	$E[Y] = \sum_{y \text{ in domain}} y \cdot p(y).$	$E[Y] = \int_{y \text{ in domain}} y \cdot f(y)\,dy.$
Functions of Random Variables (Theorems 3.2/4.4)	consider $g(Y)$ $\boxed{E[g(Y)] = \sum_y g(y) \cdot p(y).}$	$E[g(Y)] = \int_{y \text{ in domain}} g(y) f(y)\,dy.$
~~**Estimator**~~		

There are tricks that you can play with the expected value operator (see Theorems 3.3-3.5 and 4.5):

① $\mathbb{E}[a \cdot Y] = a \cdot \mathbb{E}[Y]$.

constants come out.

discrete: $\mathbb{E}[a \cdot Y)$

$= \sum \, a \cdot y \cdot p(y)$

y over domain.

$= a \cdot \boxed{\sum_y y \cdot p(y)}$

$\mathbb{E}[Y]$

② expectation is a linear operator.

$\mathbb{E}[Y_1 + Y_2] = \mathbb{E}[Y_1] + \mathbb{E}[Y_2]$.

$\mathbb{E}[Y + b] = \mathbb{E}[Y] + b$.

⚠ $\mathbb{E}[Y_1 \cdot Y_2] \neq \mathbb{E}[Y_1] \cdot \mathbb{E}[Y_2]$.

The variance (or second *central moment*) of a distribution, σ_Y^2, is

$\sigma_Y^2 = V[Y] = \mathbb{E}[(Y - \mu)^2] = \mathbb{E}[\,(Y - \mathbb{E}[Y])^2\,]$

σ_Y: standard deviation.

and it can be computed as a function of the expected values of Y and Y^2 (see Theorem 3.6):[a]

$\boxed{V[Y]} = \mathbb{E}[(Y-\mu)^2] = \mathbb{E}[\,Y^2 - 2\mu \cdot Y + \mu^2]$.

$= \mathbb{E}[Y^2] - 2\mu^2 + \mu^2 = \boxed{\mathbb{E}[Y^2] - \mu^2}$.

$\mathbb{E}[2\mu \cdot Y] = 2\mu \cdot \underbrace{\mathbb{E}[Y]}_{\mu} = 2\mu^2$.

The ~~estimator~~ for the variance is

→ **EXAMPLE.** Wackerly 7, Exercise 3.23

3.23 In a gambling game a person draws a single card from an ordinary 52-card playing deck. A person is paid \$15 for drawing a jack or a queen and \$5 for drawing a king or an ace. A person who draws any other card pays \$4. If a person plays this game, what is the expected gain?

$S = \{2, 3, 4 \ldots 10, J, Q, K, A\}$.

Let us define Y. be r.v. representing the cash payout.

possible values of Y: 15 5 −4

$p(Y=15) = \frac{8}{52} = \frac{2}{13}$.

$p(Y=5) = \frac{2}{13}$

$p(Y=-4) = \frac{9}{13}$.

$\mathbb{E}[Y] = \sum_y y \cdot p(y) = 15 \cdot \frac{2}{13} + 5 \cdot \frac{2}{13} + (-4) \cdot \frac{9}{13} = \frac{4}{13}$.

[a]You will not have to recreate this derivation on a test...but future statisticians among you should internalize it.

How do translation and scaling affect the mean and variance of a distribution?

① translation: $\mathbb{E}[Y+b] = \mathbb{E}[Y]+b$.

$V[Y+b] = V[Y]$.

$$V[Y+b] = \mathbb{E}\left[\left((Y+b)-\mathbb{E}[Y+b]\right)^2\right] = \mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right] = V[Y]$$

② Scaling: $\mathbb{E}[a\cdot Y] = a\cdot\mathbb{E}[Y]$.

$\boxed{V[a\cdot Y] = a^2\cdot V[Y]}$.

$$V[a\cdot Y] = \mathbb{E}[(a\cdot Y)^2] - (\mathbb{E}[a\cdot Y])^2 = \mathbb{E}(a^2Y^2) - (a\cdot\mathbb{E}[Y])^2 = a^2\,\mathbb{E}[Y^2] - a^2(\mathbb{E}[Y])^2 = a^2\cdot V[Y].$$

→ **EXAMPLE.** Wackerly 7, Exercise 4.33(a,c)

4.33 The pH of water samples from a specific lake is a random variable Y with probability density function given by

$$f(y) = \begin{cases} (3/8)(7-y)^2, & 5 \le y \le 7, \\ 0, & \text{elsewhere.} \end{cases}$$

a Find $E(Y)$ and $V(Y)$.

b Find an interval shorter than (5, 7) in which at least three-fourths of the pH measurements must lie.

c Would you expect to see a pH measurement below 5.5 very often? Why?

a) $\mathbb{E}[Y] = \int_5^7 y\cdot f(y)\,dy = \int_5^7 y\cdot\frac{3}{8}(7-y)^2\,dy$.

offline.

$= \frac{3}{8}\cdot\int_5^7 y\cdot(49-14y+y^2)\,dy = \cdots = 5.5$

$V[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \int_5^7 y^2\cdot f(y)\,dy - 5.5^2$

Short-cut.

offline

$= \int_5^7 y^2\cdot\frac{3}{8}(7-y)^2\,dy - 5.5^2 = \cdots = 0.15$

offline

c) $P(Y<5.5) = \int_5^{5.5} f(y)\,dy = \int_5^{5.5}\frac{3}{8}(7-y)^2\,dy = \cdots = 0.578$.

yes, $Y<5.5$ is commonly observed

Before we conclude this set of notes we will introduce Tchebysheff's theorem. Let Y be a random variable with mean μ and *finite* variance σ^2. Then, for any constant $k > 0$,

either $Y \ge \mu+k\sigma$ or $Y \le \mu-k\sigma$ $\Leftarrow$ $|Y-\mu| \ge k\sigma$

$$P[\mu-k\sigma < Y < \mu+k\sigma] \ge 1-\frac{1}{k^2}, \text{ or}$$

$$P[Y \le \mu-k\sigma \cup Y \ge \mu+k\sigma] = P[Y \le \mu-k\sigma] + P[Y \ge \mu+k\sigma] \le \frac{1}{k^2}.$$

Why is Tchebysheff's theorem important?

let $k=2$.

$\boxed{P(\mu-2\sigma < Y < \mu+2\sigma)} \ge \frac{3}{4}$.

$\Downarrow$ $\int_{\mu-2\sigma}^{\mu+2\sigma} f(y)\,dy$.

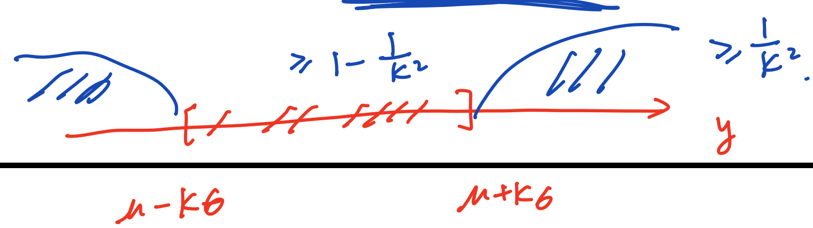
To compute $P(a\le Y\le b)$ accurately.

one needs to know pmf / pdf.

If all you know $\mathbb{E}[Y]$, $V[Y]$,

Tchebysheff's result provide

a control on $P(a\le Y\le b)$ for any distribution!



→ **EXAMPLE.** Wackerly 7, Exercise 3.167

3.167 Let Y be a random variable with mean 11 and variance 9. Using Tchebysheff's theorem, find

a a lower bound for $P(6 < Y < 16)$.

b the value of C such that $P(|Y - 11| \geq C) \leq .09$.

a) $\mu = \mathbb{E}[Y] = 11$

$\sigma = \sqrt{V[Y]} = 3.$

$P(6 < Y < 16)$

Tchebysheff's theorem:

$P(\mu - k\sigma \leq Y \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$

$$\begin{cases} 6 = \mu - k\sigma = 11 - k \cdot 3 \\ 16 = \mu + k\sigma = 11 + k \cdot 3 \end{cases} \Rightarrow k = \frac{5}{3}.$$

$$P(6 < Y < 16) \geq 1 - \frac{1}{(\frac{5}{3})^2} = \frac{16}{25}.$$

b). Goal is to find c. such that

$P(|Y - 11| \geq c) \leq 0.09.$

part 2 of Tchebysheff's thm.

$P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$

Set $\frac{1}{k^2} = 0.09.$

$k = \frac{1}{\sqrt{0.09}}.$

$C = k\sigma = \frac{1}{\sqrt{0.09}} \cdot 3 = 10$

→ **EXAMPLE.** Wackerly 7, Exercise 4.33(b)

4.33 The pH of water samples from a specific lake is a random variable Y with probability density function given by

$$f(y) = \begin{cases} (3/8)(7-y)^2, & 5 \leq y \leq 7, \\ 0, & \text{elsewhere}. \end{cases}$$

5.5 0.15

a Find $E(Y)$ and $V(Y)$.

b Find an interval shorter than (5, 7) in which at least three-fourths of the pH measurements must lie.

c Would you expect to see a pH measurement below 5.5 very often? Why?

b) Goal: find (a, b) $\underset{\text{is contained in}}{\downarrow}$ $(5, 7)$ where $P(a \leq Y \leq b) \geq \frac{3}{4}.$

Tchebysheff's theorem.

$P(\mu - k\sigma \leq Y \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$

Set $\frac{3}{4} = 1 - \frac{1}{k^2} \Rightarrow k = 2.$

$a = \mu - k\sigma = 5.5 - 2 \cdot \sqrt{0.15} = 4.725$

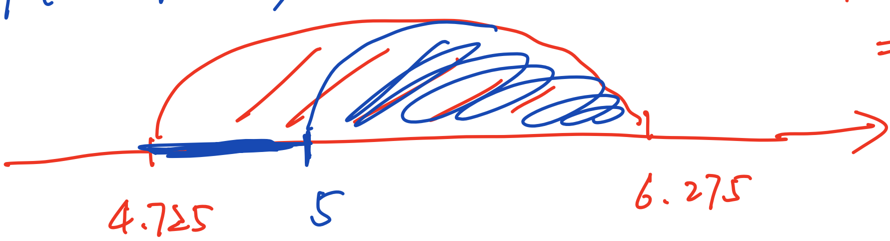
$b = \mu + k\sigma = 5.5 + 2\sqrt{0.15} = 6.275.$

(a, b)

$$\begin{cases} P(4.725 \leq Y \leq 6.275) \geq \frac{3}{4}. \\ P(Y \leq 5) = 0 \end{cases} \Rightarrow P(5 \leq Y \leq 6.275) \geq \frac{3}{4}.$$

interval

$\Rightarrow [5, 6.275]$



Review

cdf.

$F(y) = \sum_{z \le y} p(z)$ $\qquad$ $F(y) = \int_{z \le y} f(z)\,dz$

Discrete r.v
pmf $\quad p(y)$.

continuous r.v.
pdf. $\quad f(y)$.

random variable : $S \to \mathbb{R}^n$.
univariate r.v: $S \to \mathbb{R}^1 = \mathbb{R}$.

properties.

$0 \le p(y) \le 1$
$f(y) \ge 0$
$\sum_D p(y) = 1$
$\int_D f(y)\,dy = 1$

expectation.
μ, $\mathbb{E}[Y]$.

$\sum_D y\,p(y)$ $\quad$ $\int_D y f(y)\,dy$

Variance.
σ_Y^2, $V[Y]$
$\| $
$E[Y^2] - (E[Y])^2$

Tchebysheff's theorem

$P(a \le Y \le b)$

lower bound.